2. Introduction to Catalogue Interoperability

This section provides an overview of key issues in catalogue interoperability. The following topics are addressed:

- Purpose and scope of catalogue interoperability
- Catalogue Interoperability Concepts
- CIP Concepts
  - Collections data model
  - CIP as a Z39.50 profile
  - Browse data in CIP
  - Product ordering and security
- Levels of Compliance to CIP

2.1 Purpose and Scope of Catalogue Interoperability

The Committee on Earth Observation Satellites (CEOS) is comprised of international space agencies. CEOS promotes the interoperability of space agency catalogues through the definition and development of interoperability concepts. By enhancing the standardisation of EO data and information management services, CEOS enables the catalogue services to be more accessible and usable to data providers and data users world wide. EO catalogues services, as defined by CEOS, are as follows:

- search and retrieval of information about EO data products
- order of EO data products
- search and retrieval of Guide documents that complement the EO data products

Catalogue interoperability also extends beyond just the members of CEOS in promoting data access within a wider community of EO data providers and eventually to non EO data providers.

2.2 Catalogue Interoperability Concepts

The CEOS Protocol Task Team (PTT) design approach considers catalogue interoperability as the loose coupling of a federation of existing catalogue systems using a set of common protocols. The approach provides users the services available at all sites regardless of which site the user established a connection with.

- The Catalogue Interoperability Protocol (CIP) standardises the services needed for interaction between users and catalogues of EO data products.
- The ICS Guide Protocol (IGP) standardises the services needed for a user to discover EO Related documents (i.e., guide documents).
- The Interoperable Catalogue System (ICS) is a design that uses CIP and IGP as the common protocols between data providers and users of the data.

The objective of implementing CIP and ICS is to provide more users with access to more data more easily. The ICS domain can be seen in Figure 2-1 as divided into two virtual domains;

- 'CIP domain' within which CIP messages, consisting of requests and responses, are exchanged between architectural elements.
- 'IGP domain' within which IGP messages, consisting of requests and responses, are exchanged between architectural elements.

These two domains are separable and can exist independently from each other. To allow access to both domains, an ICS client was designed with a CIP Client component and an IGP Client Component. To enable consistency in the ICS domain a Collections Management Tool (CMT) was defined to update the contents of both the CIP and IGP domains simultaneously.

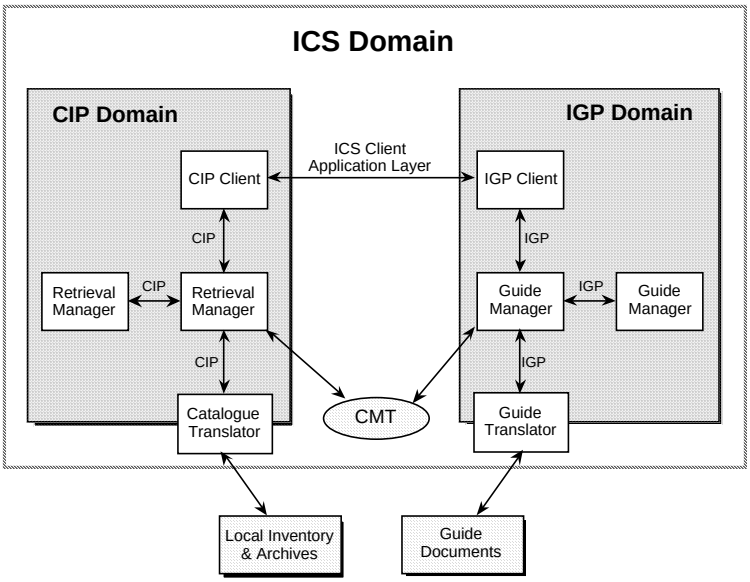

Figure 2-1: ICS Domain

To support transparent access to multiple catalogues, a three tier structure was used to design the ICS space.

Clients exchange messages with a middleware layer which in turn interacts with multiple catalogue servers. The middleware provides the routing and translation services to allow client requests to be presented at the multiple heterogeneous catalogues. The middleware is of two types of elements: Managers and Translators.

- Managers provide an access point for clients and route the requests to the various servers.
- Translators, bound with the clients and servers, translate CIP or IGP to and from the native protocol of the client or server. Future client and server developments may use CIP or IGP directly and hence not require translators

This approach supports a diversity of clients, and servers. Clients may be used directly by a human user or may be an agency system acting on behalf of a user. Depending on the design of an existing catalogue system, services may be provided by different servers. The ICS approach allows the use of either one or more translators to access inventory, browse, ordering, and user profile services that reside at a specific site.

This architecture is also applicable for small data providers, such as university research groups, who are unable to provide an adequate manager functionality at their site but still wish to join the ICS domain. Their local inventory entries or/and guide documents can be made available to the ICS community by providing a link from their catalogue server to another agency's manager.

The CIP Space provides for the loosest coupling needed to achieve catalogue interoperability among a wide community of EO data providers. A range of design solutions is permitted by the CIP. To provide for a higher degree of uniform services at the cost of additional agreements between agencies, additional design criteria for interoperability are defined in the ICS design document. The difference between CIP Space and the ICS is depicted in Figure 2-2.

The CIP Space is defined by all federations and organisations that provide catalogue services using the CIP. Those CEOS agencies which provide services, communications and systems management compatible with the ICS design make up the ICS. Its should be noted that while all ICS members must implement CIP, guide handling is considered an optional element of ICS and an ICS members may choose not to implement IGP. Note that federations other than CEOS may choose to use the ICS design as the basis for their federation.

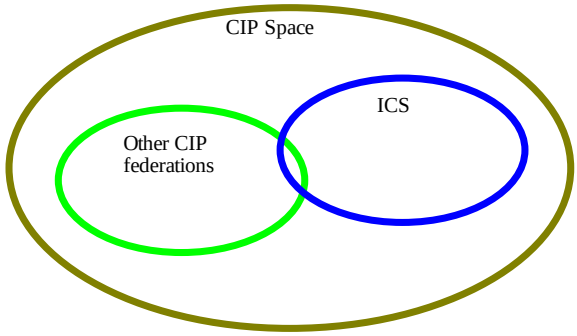

Figure 2-2: VENN diagram of CIP Space and ICS

Assuming query and result routing between geographically dispersed sites (see Figure 2-1), an agreed middleware layer and its interfaces to users and providers needs to be in place. To define such a system, the PTT have established the following ICS standards:

CIP/ICS:

- Search Standardisation - the functions and procedures how search can be invoked and executed both by end user clients and middleware; the search language and it's syntactical rules;
- Retrieval Standardisation - the procedure for retrieval of query results and other objects which need to be presented to the client; an exact description of what is the retrieval content, its format and meaning.
- Attribute Standardisation - the list of fields or attributes which can be searched and the definition of their semantic meaning. In addition CIP provides a local attribute mechanism which allows the data producer to extend the set of attributes to include more information to assist the user in the selection of EO products of interest.
- Dynamic Client Configuration Standardisation - a mechanism which is used for dynamic client configuration, i.e. to make a client understand which functionality can be invoked at the server side, which attributes are understood, and what are their meanings.
- Order Procedures Standardisation - a common way of defining product lists and associated order, packaging and online/offline delivery options; communication of price and accounting information; order and delivery addresses.
- Security Policies Standardisation - a reliable and secure mechanism for user authentication and authorisation.

IGP/ICS:

- Guide Document Discovery and Retrieval Standardisation - a mechanism which allows full text or fielded searches on documents relating to EO data holdings. This mechanism uses HTTP protocols and technology to enable the use of standard web browsers and discovery services such as Alta Vista to discover and retrieve EO related documents of interest. This service provides the general WWW community access to documents describing EO data. This service also provides pointers to EO data related to each document.

2.3 CIP Concepts

2.3.1 Collections Data Model

In an interoperable catalogue environment it is essential to organise metadata by distinguishing user and provider views as well as archive-oriented and theme-oriented structures. It is important to define a mapping between the different views and structures which often will lead to a hierarchical relationship of collections. The CIP data model is based on the notion of collections. A collection may contain descriptors for data products (product descriptors, also known as inventory entries), or descriptors for other collections (collection descriptors). In addition to the value of collections for presentation of data organisation to users, collections provide the mechanism for routing distributed searches. When a collection contains both local and remote members, the Retrieval Manager may search the local site as well as sending the search on to the remote site.

The collection concept is visualised in Figure 2-3 below. The collections in the diagram are numbered so that their relationship can be easily seen; they do not represent the naming of collections in an actual implementation. The terminal collections (labelled '1.x') group the product descriptors as is appropriate. As can be seen the collections can overlap each other and product descriptors can appear in more than one collection. Above the terminal level collections, there are non-terminal collections that group together any number of other collections. The grouped collections do not all have to be at the same hierarchical level and this grouping of collections can continue to any hierarchical level, with existing collections being included at any other arbitrary level. A non-terminal collection could group together terminal collections and other non-terminal collections (as the link between collections 3.1 and 1.5 shows). Also, a terminal collection could exist without a relationship to a higher collection (i.e. collection 1.9), or a non-terminal collection could exist with no relationship to lower collections, in other words a collection without members (i.e. collection 2.5). Collection 1.9 can not be reached by a hierarchical search, but could be located if its URL was made public (an example of such a collection may be a persistent result set or a collection under construction).

Collections can be used to group data together which have a similar semantic theme. All collections support the search mechanisms defined in the CIP. The CIP defines two types of searches which a CIP user may request:

- **Collection Search**: finds collections of interest without searching collections containing products
- **Product Search**: finds individual product descriptors which may eventually lead to the order of an actual product.

Additionally, the user may request that the search be contained locally to the target Retrieval Manager (i.e., a local search), or request that the search be propagated to other Retrieval Managers based on the collections (i.e., a distributed search).

Two types of collections are defined:

- **Registered collections** are owned by the EO data provider and described by a collection descriptor. These collections support the full range of CIP access services including discovery, navigation, location and searching. There are two types of registered collections which are distinguished by the purpose for which the were created:
  - **Archive collections**: this type of collection is likely to be created by data providers to organise their archives and facilitate access to the product descriptors (i.e. analogous to an inventory containing inventory entries)
  - **Theme collections**: this type of collection may be set up by data providers or users who want to organise some of their data into groupings which differ from their provider archive collections (i.e. from the baseline inventory), for the convenience of their users, for example, based on the geographical area covered, the scientific discipline supported by the data, the instrument type, etc.
- **Unregistered collections** are likely to be created and owned by an end user of EO products that has created and populated the collection to obtain a single source of thematic information. This will then enable further analysis or easy access by themselves or other users. These are collections of potentially quite disparate item descriptors of interest to a relatively small user community researching a particular theme, i.e. in the example, the mid-west flood of 1993. Unregistered collections need not be described by a collection descriptor and their collection attributes cannot be discovered in a search.

An unregistered collection may be registered by an ICS data provider. The process of registration may vary significantly between ICS data providers, but minimally requires the creation of a collection descriptor and the insertion of that collection descriptor into the ICS data providers collection structure. Generally the registration of a collection will also involve a scientific review and the transfer of ownership of the collection from the end user to the ICS data provider to ensure accuracy and long-term availability of the registered collection. Further description of the registration process can be found in the ICS Collections Manual [CM].

Note that these category definitions are not mandatory for the CIP to operate, but they aid the discussion about collections. The CIP does not distinguish between the categories (the Retrieval Manager does, however) and the same CIP search and retrieval services are applicable to all collections. The Retrieval Manager does make use of collections for routing of distributed queries. Standardisation of collection definitions is provided as part of the ICS design.

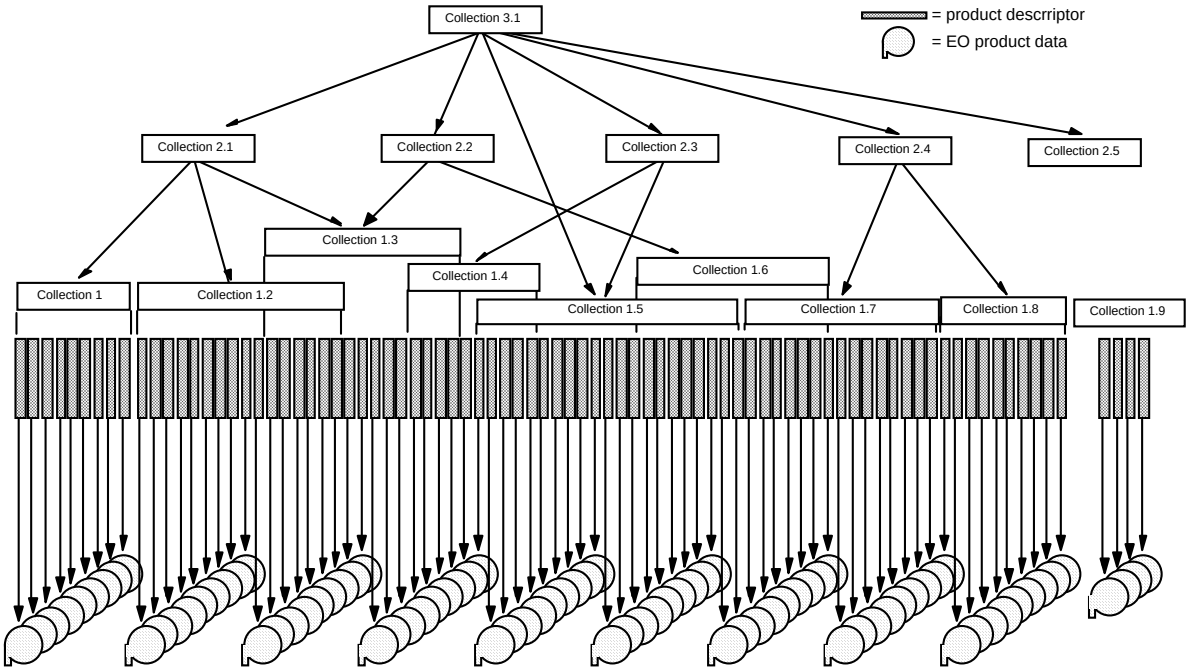

Figure 2-3: The Concept of A Collection

2.3.2 CIP as a Z39.50 Profile

Based on a set of user requirements and an analysis of existing communication standards, Z39.50 [Z39.50] was selected as the base protocol for CIP. The Z39.50 protocol is designed for information search and retrieval within a generic domain which, together with the powerful services and data structures it supports, makes it an ideal basis of an EO domain search and retrieval protocol.

CIP has exploited and extended the services of Z39.50 to provide distributed searching, extensions to attribute set definitions, and the definition of a secure ordering service. CIP is a profile of Z39.50, i.e. it defines the use of the Z39.50 facilities within the CIP domain and defines the attributes that are used to search and present EO information. CIP extends Z39.50 for distributed searching by supporting the collection data model discussed in section 2.3.1 which allows hierarchies of related collections to be constructed and searched. Other Z39.50 profiles include GILS, GEO and the Digital Collections Profiles.

The GEO Profile [GEO] supports Geographic Information Systems (GIS) applications and thus is of special interest to users of EO data. For this reason an alignment of the CIP and GEO profiles was made. The objective of this alignment was to allow both GEO and CIP clients to search and retrieve records from databases defined by either profile, and thereby maximise interoperability. The alignment was helped by the similarity of the spatial and temporal attributes of the metadata, but needed to take into account the different data models in CIP and GEO. It should be emphasised that the CIP/GEO interoperability is for search on the intersection of CIP and GEO attributes and HTML formatted records are retrieved. There is no interoperability on the more advanced functions of CIP such as ordering and security.

Additional support for compatibility is provided by the requirement that Retrieval Managers must support access by any Z39.50 Version 2/3 compatible client.

2.3.3 Browse Data and CIP

The main aim of Browse data is to help users in evaluating EO products. Browse data are typically reduced resolution or summary data derived from the EO product data itself. They are delivered to the user dynamically over the network during a user query session. It is important to note that although most catalogue systems will provide some form of browse data retrieval, it is not a mandatory CIP service. The form and content of browse data is dependent on the nature of the associated EO data and the data selection criteria necessary for a science discipline to evaluate the EO data. Browse data in the CIP is seen as one of the following forms:

- Browse compound - compound attribute containing attributes describing the Browse data, including the temporal and spatial coverage of the browse data.
- Browse data - attribute that either contains the browse data or provides a reference to it (browse pointer).

2.3.4 Product Ordering and Security

CIP supports a wide range of ordering services, e.g. the specification of order options or mechanisms for authentication and non-repudiation of orders. A user can retrieve the order options associated with a product where the order options may be any combination of processing, scene selection and delivery options.

CIP allows a local order handling system to define or refer to its own local order options. But the CIP also contains a standard mechanism to define order options.

A user can request a quote for a specific order and submit the order. The order process is monitored by the Retrieval Manager and can be queried later by the user to determine the status of the order. To support ordering of data for which a user must have privileges or for orders which the user will be charged, a authentication scheme has been defined. The authentication supports digital signatures using either a shared (symmetric) key approach or an public (asymmetric) key approach. Authentication allows the Retrieval Manager to identify the user with an appropriate level of confidence and enables the Retrieval Manager to log the authenticated user requests to provide non-repudiation. The CIP security approach avoids the need to transfer password information over the network. Future enhancements to CIP anticipate the ability to support the transfer of financial information to support billing.

If a provider decides to create a collection containing browse products, a user would be able to order browse products via the CIP. In this case the browse product would be considered as any other data product and not any longer as ‘browse’ as defined in the previous section.

In addition to section 3.5.8 of this document, the document ‘Order Options Amendment to the CIP Specification’ [ORD] provides other important information on the CIP supported ordering services.

2.4 Levels of Compliance to CIP

The ICS SDD and CIP specification are detailed documents with many services and mechanisms specified. As discussed earlier in this section, agencies may choose to implement a wide range of these services in their CIP Clients and Retrieval Managers. It is critical for the designers and implementers of these software components to understand what capabilities are critical to the minimal operations of the ICS and must be implemented in all components versus those capabilities which are optional. In addition it is assumed that various CIP based components will be available either as shareware or commercial software. The developers of ICS or other CIP federations will need a method to categorise and select among these available components. For this reason, compliance levels have been defined within both the ICS SDD and the CIP Specification.

- *ICS Compliance* (in Section 9 of the SDD) discusses the minimum requirements a site must meet to become a member/node of the ICS federation (RM and underlying catalogues).
- *CIP Compliance* (Annex H of this document) discusses the specific CIP messages and parameters that must be supported by a CIP Client or an RM.

These compliance concepts are interdependent since in order to support a specific ICS service, the RM and CIP Client must support the CIP messages which enable that service.